# Deep learning for robotic strawberry harvesting

1st Xiaodong Li
*School of Computer Science*
*University of Lincoln*
Lincoln, United Kingdom
doli@lincoln.ac.uk

2nd Charles Fox
*School of Computer Science*
*University of Lincoln*
Lincoln, United Kingdom
scoutts@lincoln.ac.uk

3rd Shaun Coutts
*Lincoln Institute of Agri-Food Technology*
*University of Lincoln*
Lincoln, United Kingdom
scoutts@lincoln.ac.uk

**https://doi.org/10.31256/Bj3Kl5B**

***Abstract*—We develop a novel machine learning based robotic strawberry harvesting system for fruit counting, sizing/weighting, and yield prediction.**

***Index Terms*—machine vision, cascade detector, Harr/LBP feature, yolo.**

## I. Introduction

Strawberries are a high-value crop all around world, but during harvest season, due to the weather fluctuations, strawberry yields can vary greatly on a daily basis. It remains a challenge for farmers to efficiently manage labour and transport which rely heavily on accurate prediction of near future production.

Traditional yields estimation is time consuming and labour intensive. With the maturation of low cost camera sensing and corresponding vision processing technology, machine vision has become a potential alternative to traditional way. It has high adaptivity to variant image quality.

The Viola-Jones cascade detector used here is well supported in the OpenCV library with both Haar-like feature [1] and Local Binary Patterns(LBP) feature [7] [6]. Notably it was the first real-time (CPU-based) face detector and recognizer [1].

The project was initially low cost CPU-based platform which was later upgraded with an Nvidia GPU card (GTX TITAN X). The required expanding and enhancement on existing functionality, e.g, ripeness prediction, brought our attention to Deep learning.

Althugh there are already success [3] with deep learning, this quick development integrated with existing system is notable. By passing the results of the Viola-Jones detector to a YOLO (You Only Look Once) [2] [4] based deep learning system, the classification is achieved in various color category and improvement on detection accuracy, etc.

## II. System Architecture and Methodology

### A. System design

The system is shown in Fig.1, where "Video input" of 'back2back' opposite facing 2-camera provides consecutive image frame (fps: 30); "Image Tracking" is to track individual images simply by a template matching strategy to resolve overlapping, and it reduces the computational cost by operation conducted only on parsed none overlapped image, which improved the running speed for detectors as detection doesn't need apply on each frame, this makes real time detection possible and overcomes double counting; "Fruits Detection" applies a trained cascade detector for strawberry detection. The cascade detector model has a single class for the full range of strawberry including flowers; To further split this class into different categories, the "Classification" used a trained YOLO model to classify color difference, i.e. green, white and red, including a leaf model to further remove false detection; "Count/Size/weight" can provide different metrics for different purpose, e.g the number of strawberries in different color, with a pre-defined fruit shape model (e.g sphere, cylinder and cubic, etc) the metrics such as diameter, circumference, or volume can be obtained, further combined with strawberry's density, the overall harvest weights can be estimated, etc. currently, a simple method based on the average distance between the camera and the strawberry rack is applied, which combines the calibrated camera parameters to calculate the dimension or volume according to the strawberry shapes; according to weather condition and farmer's experience, the "Yields Prediction" provides guide on the amount of ripe strawberries for the harvest in different time, etc.

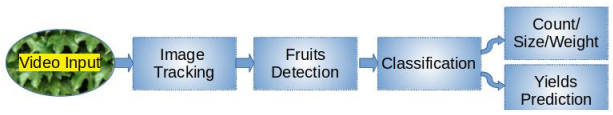



Fig. 1: System Structure

### B. Viola-Jone Cascade Detector

Viola-Jones [1] detector is a Harr-like cascade detector, which can reduce the computational cost on the image intensities. Using a sliding window over the image, Harr-like features are calculated and compared by the difference from a learned threshold. While single features are weak learners, a classifier cascade is used to obtain stronger classifiers by combining them. Viola-Jones takes a variant of adaptive boosting (AdaBoost) for feature selection and training a detector using object and background images. A single classifier consists of a weighted sum of many weak classifiers, where each weak classifier is a threshold on a single Haar-like rectangular feature. The weight associated with a given sample is adjusted based on whether or not the weak classifier correctly classifies the sample. Haar-like features boost real-time detection for human faces, but still infeasible for larger image.

*1) Local Binary Patterns features:* LBP features [7] [6] utilized here provide a suitable alternative. It is often a powerful (speed) feature for texture classification. A LBP vector can be simply calculated in an image cell (e.g 16x16 pixels of sub-window), where the pixel is compared to each of its 8 neighbors along a circle. The pixel's value is the concatenation of a binary "0" and "1", which is assigned by comparing (less or greater) with each neighbor. This 8-digit binary number is usually converted to decimal for convenience. The histogram over the cell for the frequency of each "number" occurring is computed and regarded as a 256-dimensional feature vector.

### C. Deep Learning

YOLO [2] [4] is one of the most effective accurate real time object detection algorithms. It applies a single neural network to the full image, and then divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by their predicted probabilities, which output the recognized object after non-max suppression applied.

Because of the small size of strawberry image (24x24 pixels) in Viola-Jones, it can not provide enough information for precise classification with Haar/LBP feature. We then utilized a self-trained YOLO model for color-based object classification to predict the ripeness of strawberries. Without manual relabeling the existing datasets, an effective and innovative integration of state-of-the-art advanced YOLO with existing system is achieved successfully.

### D. Data Sets and Annotation for YOLO

The initial labeled training image datasets (45K) were collected across the world e.g Fig.2 and cropped ones like Fig.3. Inspired by YOLO training mechanism, without standard YOLO labeling, the innovative auto-'one4one' self annotation method on existing data is shown below:

$center - x = x/w(0.5); center - y = y/h(0.5); width = w/W(1.0); height = h/H(1.0);$

$x$: $x$-coordinate of center of the bounding box;
$y$: $y$-coordinate of center of the bounding box;
$w$: width of the bounding box;
$h$: height of the bounding box;
$W$: width of the whole image;
$H$: height of the whole image;

(a) Indoor (b) Outdoor

Fig. 2: Typical Strawberry Image

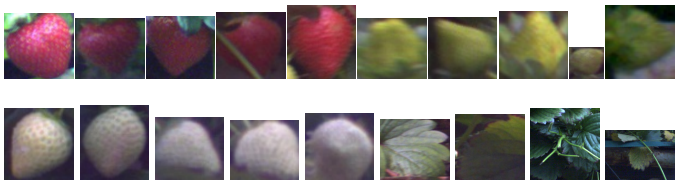


Fig. 3: Strawberry Training Images

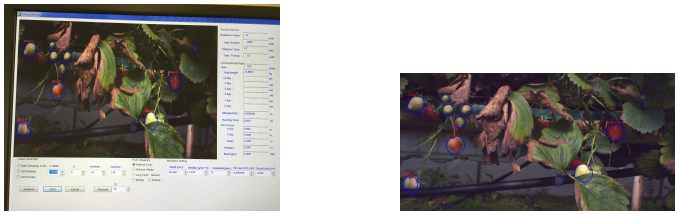
## III. Results and Summary

System was developed with C/C++, and the snapshots in Fig.4 depict system configuration&monitoring(a), detection image(b). In (a), the metrics(unit), which is configurable as request, for sizing and weighting are diameter(mm) and weights(gram) respectively.

The classification examples in Fig.5 are red strawberry (a), green one (b) and Leaves (c).

The system was tested on a 80m strawberry growing rack, through the absolute difference of the numbers between manual and machine counts divided by manual ones, we have overall (all categories together) accuracy of about 90% on counting/detection, 95% on classification for red strawberries, and 80% for green and white color respectively with the model trained on 300 samples for each category, which can be further improved by more carefully selected samples added for training.

The sizing has a threshold above 15mm in diameter, together with other functionality, we believe the weight prediction can be improved with more accurate assumed depth measurements, shape model (cylinder used here) and density setting, etc.



(a) System GUI (b) Detection Outputs

Fig. 4: System Features

(a) Red Strawb (b) Green Strawb (c) Leaves

Fig. 5: System Classification

## REFERENCES


[1] Paul Viola , Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", IJCV 2001.

[2] Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement", Computer Vision and Pattern Recognition, April, 2018.

[3] M. Mahmud, M. S. Kaiser, A. Hussain and S. Vassanell, "Applications of Deep Learning and Reinforcement Learning to Biological Data", IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 6, pp. 2063-2079, June 2018.

[4] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger", Computer Vision and Pattern Recognition, December, 2016.

[5] N. Cristianini , J. Shawe-Taylor, "An introduction to support Vector Machines: and other kernel-based learning methods," , Cambridge University Press, 2000.

[6] L. Wang , DC. He, "Texture Classification Using Texture Spectrum,"IEEE Transactions on Pattern Recognition,vol. 23, pp.905–910, 1990.

[7] DC. He , L. Wang, "Texture Unit, Texture Spectrum, And Texture Analysis," IEEE Transactions on Geoscience and Remote Sensing,vol. 28, pp.509–512, 1990.